

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
25 July 2002 (25.07.2002)

PCT

(10) International Publication Number  
WO 02/057957 A1

(51) International Patent Classification: G06F 17/30

(21) International Application Number: PCT/US02/01429

(22) International Filing Date: 16 January 2002 (16.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/262,050 16 January 2001 (16.01.2001) US

(71) Applicant: SANGATE SYSTEMS, INC. [US/US]; 144  
Turnpike Road, Southborough, MA 01772 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

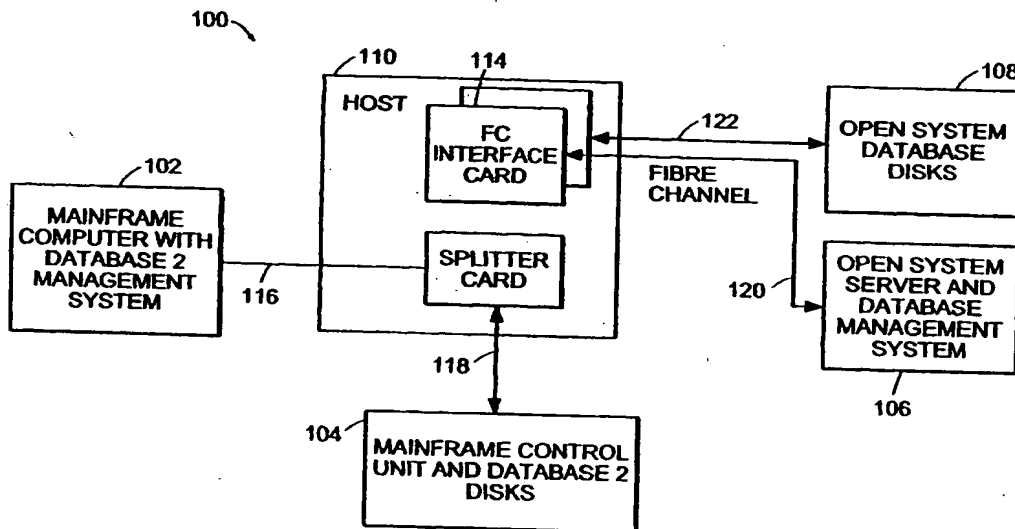
— with international search report

(72) Inventor: WINOKUR, Alexander; Hatishbi Street 108a,  
34521 Haifa (IL).

(74) Agents: KENNARD, Wayne, M. et al.; Hale and Dorr  
LLP, 60 State Street, Boston, MA 02109 (US).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SYSTEM AND METHOD FOR CROSS-PLATFORM UPDATE PROPAGATION



(57) Abstract: A system and method is described for use in the propagation of cross-platform updates of databases in, for example, an Enterprise system. Given that these platforms may operating using different operating systems and database software, the system and method are not application dependent. The system and method effect the cross-platform propagation of updates through the use of I/O streams. The system and method include a host (110) which supports an interface (114) and a fiber channel interface card (122) to handle the connectivity to the open system server DBSM (116) and the open system database disk (108). Accordingly, the system and method do not use valuable and expensive CPU cycles nor excessively load the TCP/IP communication lines to perform the desired cross-platform propagation of updates for platform databases.

## **System and Method for Cross-Platform Update Propagation**

### **Cross Reference to Related Applications**

- 5        This application claims priority to provisional application no. 60/262,050, entitled "A Method and an Apparatus for Cross DB Update Propagation," filed on January 16, 2001.

### **Field Of the Invention**

- 10       The present invention relates to Enterprise systems that have front-end and backend data repositories that require updating. More specifically, the present invention relates to Enterprise systems that have front-end and backend data repositories that require the propagation of cross platform updates.

### **Background of the Invention**

- 15       The wide use of the Internet has provided a new mode of business which has been frequently referred to as "e-Business" or "e-Commerce." The heart of e-Business is the ability of system users, which at times are customers to be able to perform a variety of transactions using a web browser or other device that allows them to connect to the Internet. In the context of an "Enterprise," "Enterprise Data" may now be accessed and updated by customers in a much less controlled environment. In
- 20       Enterprises that are engaged in e-Business where, in fact, customers do have such access, there is a trend to isolate the critical backend data repositories from the front-end data repositories. This is felt necessary to prevent the backend data repositories from being manipulated by customers through web browsers. Further, the front-end and backend data repositories may be different types of systems running different
- 25       types of database software. For example, the front-end repositories may be S/390 systems from IBM running Database2 software from IBM, Inc., while the backend data repositories may Unix/NT based systems from Sun Microsystem Inc./Microsoft, Inc. running Oracle database software.

- 30       Given that front-end and backend data repositories are all part of the same Enterprise system, it is very likely that some of the updates to the backend data repositories will propagate to the front-end data repositories and vice-versa. In order

to carryout such cross-platform updates, normally, there is very specialized software components added to the respective platform software to effect the desired update propagation. Generally, platform-to-platform propagation using this software is carried out using the system's Transport Control Protocol/Interface Protocol ("TCP/IP") communications lines.

The add-on software components are expensive to the system in a variety of ways. These expenses include the amount of CPU cycles needed from both the front-end and backend servers, the loading of the TCP/IP communications lines, and there must be specially developed software for each specific enterprise logic. This latter requirement also means that additional supporting cross-system locking mechanisms and logic must be developed. As can be suspected, with the need to handle these issues there is a strong likelihood that applications developed in this environment will have bugs and be susceptible to various types of failures.

As an example, in any two platform system, which includes Enterprise systems with front-end and backend data repositories, there can be real-time consistent update propagation by including transaction application logic to the respective database servers. In the case of the cross platform propagation of updates, one of the transaction applications will be the source transaction application and the other will be the destination transaction application. Each of the transaction application is capable of serving as a source or destination transaction application. Therefore, each is developed so that it is capable of propagating updates that it intercepts to a program location at the destination database of the other platform. The program at the destination database will acquire and retain the appropriate records relating to the updates propagated to the destination database. The destination database will then apply the updates to the destination database.

After the successful update of the destination database, the destination program will transmit a "successful completion status" message to the source transaction application. This message will then permit the sources transaction application to complete its internal processing for the propagation. The system that has just been described, however, uses the respective server CPU cycle times and loads the TCP/IP communications lines to effect update propagation which is undesirable.

There is a need for a system and method the will more efficiently and effectively perform cross-platform update propagation without the problems and expense found in prior systems.

### Summary of the Invention

5       The system and method of the present invention will effect the efficient and effective cross-platform updating of databases which is not application dependent. The system and method of the present invention do not require additional cross-system locking mechanisms, uses a very limited number of server CPU cycles, and do not use the TCP/IP communications lines.

10       The system and method of the present invention may be carried out by updates for the respective databases being intercepted as reflected in their respective database logs. The intercepted updates are then applied to the databases by directly writing them to the appropriate database disks. According to the present invention, the interception of the updates and the writing of them to the disks are carried out using  
15 the I/O streams, thereby bypassing the need to use the respective different server CPUs and the TCP/IP communications lines. As an example, the I/O streams include Enterprise System Connection ("ESCON"), Small Computer System Interface ("SCSI"), or Fibre Channel streams.

20       An object of the present invention is to provide a system and method to effectively and efficiently propagate cross-platform updates between two discrete databases systems.

Another object of the present invention is to provide a system and method that will effect cross-platform update propagation without the need to use platform server resources or TCP/IP communication lines.

25       A further object of the present invention is to provide a system and method that will effect the propagation of cross-platform updates between two discrete database systems at the I/O streams associated with a system that may include the two discrete database systems.

30       These and other objects will be explained in greater detail in the remainder of the specification, and in light of the drawings and the appended claims.

## Brief Description of the Drawings

Figure 1 shows a representative system incorporating the system and method of the present invention.

Figure 2 shows a detailed representative view of a splitter that is used in implementing the present invention.

Figure 3 shows detailed view of the common I/O buffer that is an element of the representative splitter shown in Figure 2.

## Detailed Description of the Invention

The system and method of the present invention are directed to the propagation of cross-platform updates of databases in, for example, an Enterprise system. Given that these platforms may be operating using different operating systems and database software, the present invention is not application dependent. The present invention effects the cross-platform propagation of updates through the use of I/O streams. Accordingly, the system and method of the present invention do not use valuable and expensive CPU cycles nor excessively load the TCP/IP communication lines to perform the desired cross-platform propagation of updates.

An overview of the system and method of the present invention is that when there is a system, such as an Enterprise system with at least two platforms, in which cross-platform database updating is desired, the specific updates for the database(s) of a particular platform is intercepted by a system element added for that purpose. The intercepted update information is then written to the log of the original database, the source database, to which the update was intended and to the log of the database for the other platform(s), the target database(s). The updates are then written to the source and target databases in a manner that does not interfere with other activity of the system.

The present invention may be implemented at least in part through a platform that is described in commonly assigned, co-pending U.S. patent application Serial no. 09/605,438, titled "Device, System, and Method of Intelligently Splitting Information in an I/O System." This application is incorporated herein by reference.

Before describing the system and method of the present invention in greater detail, there are certain terms that will be used in the description that will have the following definitions:

Notations

$R_j$	The identification of a source database record $R_j$ .
$r_j$	The identification of a target database record $r_j$ .
$R_{j,k}$	An entry in the source database log indicating the update of record $R_j$ by the transaction identified as $k$ .
$r_{j,k}$	An entry in the target database log indicating the update of record $r_j$ by the transaction identified as $k$ .

Database Transactions

$T(R_i, R_j, R_k, \text{Read}(R_i, R_m, R_n, R_o))$	Source database records $R_i, R_j$ , and $R_k$ are updated based on the values of source records $R_i, R_m, R_n$ , and $R_o$ . Therefore, each transaction $T$ has a read domain $R$ , records $R_i, R_m, R_n, R_o$ , and an update domain $U$ , records $R_i, R_j, R_k$ .
---	--

- 5 Parallel Transactions: Transactions  $T_i$  and  $T_j$  are parallel if one of the following is true:

1.	$T_i$ starts after $T_j$ starts but before $T_j$ ends.
2.	$T_j$ starts after $T_i$ starts but before $T_i$ ends.
3.	$T_i$ starts before $T_j$ starts and ends after $T_j$ ends.
4.	$T_j$ starts before $T_i$ starts and ends after $T_i$ ends.

Atomic Transactions: A transaction  $T=(R, U)$  is atomic if for any other parallel transaction  $T_i=(R_i, U_i)$ , the following is true:

1.	There are less than 2 records in $(R \cup U) \cap U_i$ . This also may be stated as the operation $O((R \cup U) \cap U_i) < 2$ (the "atomicity equation"). That is, the transaction $T_i$ cannot change more than one record in either the read or write domain of transaction $T$ .
2.	If $O((R \cup U) \cap U_i) > 1$ , then all read/update operations which violate $O((R \cup U) \cap U_i) < 2$ , occurred either before $T$ started or after $T$ ended.

10

Update Propagation: For the purposes of cross-platform propagation the following applies:

1.	DB1 and DB2 are two independent database servers running on different systems.
2.	DB1 is the source database where the updates originate and DB2 is the target database where the updates will propagate.
3.	$R_i$ may be a record that in DB1 and $r_i$ is a corresponding record based on predefined mapping in DB2.
4.	$T(R_i, R_j, R_k, \text{Read}(R_i, R_m, R_n, R_o))$ is a transaction executed by the source database, then following transaction execution, and after applying the update propagation

	records $r_i$ , $r_j$ , and $r_k$ , which correspond to $R_i$ , $R_j$ , $R_k$ , will have the same values as $R_i$ , $R_j$ , $R_k$ .
--	--

Consistent Update Propagation: This is an update that complies with the following rules:

1.	Target database transaction atomicity is not violated. That is, updating $r_i$ , $r_j$ , and $r_k$ , with values of $R_i$ , $R_j$ , and $R_k$ does not violate the atomicity rules for any of the transactions executed natively by DB2, the target database, in parallel to the update propagation.
2.	The atomicity of the transaction represented by the update propagation is not violated by any DB2, the target database, native transaction.

5 According to consistent update propagation definition, for any target transaction,  $T=(R_i, U_i)$ ; and for any source update propagation transaction,  $T=(U)$ ,  $O((R \cup U) \cap U_i) \leq 2$ . This shows that an update propagation transaction on the target database DB2 consists only of the update domain, since the update is not dependent on any target database DB2 values. Thus, for update propagation transactions,  $R=0$ .

10 Real-Time Update Propagation: An update propagation that completes execution together with the original transaction that initiated the update propagation.

Database Log: A set of primary and secondary log files consisting of log records that record all changes to a database. The database log is used to roll back changes for transactions that are not committed and to recover a database to a consistent state.

15 Figure 1, generally at 100, shows a representative system that incorporates the system and method of the present invention. The system includes mainframe computer 102 that has a database management system ("DBMS") associated with it. The DBMS will control the database system associated with the mainframe. As an example, for mainframe computer could be an IBM S/390 system.

20 The mainframe control unit and database disks are shown at 104. The mainframe and DBMS at 102 connect to mainframe control unit and database disks at 104 so that information data can be retrieved from, or added to the database disks. The DBMS and database may operate according to Database2 operating software from IBM.

25 Again referring to Figure 1, system 100 also has open system server 106 that includes a DBMS that is used for controlling the associated database system that is shown at 108. The open system server may be a SAN/Solaris server. The DBMS and the database may be an Oracle DBMS and Oracle disks, respectively.

Host 110 is the system 100 element through which mainframe/DBMS 102 connect to mainframe control unit/database disks 104, and open system server/DBMS connects to open server disks 108. Host 110 includes intelligent splitter 112 that connects mainframe/DBMS 102 connect to mainframe control unit/database disks 104. Further, host 110 has interface 114 that connects open system server/DBMS 106 to open system database disks 108. Host 110 also provides connectivity between splitter 112 and interface 114 so that propagation in accordance with the present invention can take place.

According to Figure 1, host 110 uses one port of splitter 112 to handle ESCON connectivity 116 to mainframe/DBMS 102 and second port to handle ESCON connectivity 118 to mainframe control unit/database disks 104. Host 110 also supports interface 114, which may be a fibre channel interface card. Fibre channel interface card 114 handles connectivity 120 to open system server/DBMS 106 and connectivity 122 to open system database disks 108.

The system and method of the present invention may be implemented in a system such as is shown in Figure 1. In such a system, all "writes" to the logs of the source and target databases, and all "reads" to the target database are intercepted. The interception of this information or data takes place in the I/O streams. According to aspects of the present invention, host 110 may be programmed to direct splitter 112 to intercept all write commands in the I/O stream from mainframe/DBMS 102 and reads from the I/O stream from open system disks 108 as controlled by the open system server/DBMS 106.

The I/O intercepts and I/O based activity with regard to the present invention will be performed at host 110 and splitter 112 according in commonly assigned, co-pending U.S. patent application Serial No. 09/605,493, titled "I/O System Supporting Extended Functions and Methods Thereof, " the contents of which is incorporated by reference.

A detailed, representative view of splitter 112 is shown in Figure 2, generally at 200. In Figure 2, splitter 112 includes Port A, Port B, common I/O buffer 220, local processor 230, local processor memory 240, and communications bus 250. Ports A and B communicate with external connections 210A and 210B to receive and transmit data, for example, according to ESCON protocol. Each of the Ports also communicates with common I/O buffer 220 using bus 214. It is understood that each

of the Ports has the capability of read visibility into the entire buffer and write visibility to specific buffer areas associated with the Port.

Common I/O buffer 220 is used to store frames received from an external link. It will also contain frames created by local processor 230. It is from I/O buffer 220 that update data is intercepted and inspected.

Local processor 230 runs software in memory 240 to control splitter 112. For example, local processor 230 may run software that can read and/or write states to the Ports to control operation. Further, since local processor 230 can communicate with common I/O buffer 220, programs may be run to read and/or write information to common I/O buffer 220. These latter programs, for example, may be used to inspect I/O streams.

Local bus 250 is used for communications among processor 230, common I/O buffer 220, Port A, Port B, and processor memory 240. This bus will permit interrupt, command, address, and data information to be passed among the components connected to bus 250. That is, bus 250 facilitates communications to the uniquely addressed components connected to it.

Figure 3, generally at 300, shows splitter 112 that is shown in Figure 2 with common I/O buffer 220 shown in greater detail. As is shown in Figure 3, the unique address space of common I/O buffer 220 is subdivided so that each of the Ports and processor 230 is associated with a unique sub-address space. For example, common buffer 220 may be logically divided into three equal size, non-overlapping, memory segments 220A, 220B, and 220C. Port A may be associated with segment 220A, Port B with segment 220 B, and processor 230 with segment 220C. This is only meant to be representative and other configurations of the I/O buffer are possible and still be within the scope of the present invention.

Again referring to Figure 1, there are two databases systems shown. The first is shown as mainframe control unit/database 104 and the second as open system database 108. As described, each may be running on different software. For example, mainframe control unit/database 104 may be running Database2 software, while open system database 108 may be running an entirely different type of database software, such as Oracle database software. For purposes of description only, mainframe control unit/database 104 will be referred to as DB1, the source database, where the updates originate, and open system database 108 will be referred to as DB2, the target database, where the updates are to propagate.

DB1 and DB2 each have a database log associated with it. These data structures are used in implementing the present invention. Among other things, the DB1 source database log and DB2 target database log contain records for all update transactions and their associated records. For example, for update transaction  $i$ ,  $T_i =$  Update ( $R_i, R_j, R_k, \text{Read } (R_l, R_m, R_n, R_o)$ ), the following data would be recorded in the database log:  $BT_i$ , the begin transaction; the list of updated records with the updated new values,  $R_{li} = V_i, R_{ji} = V_j, R_{ki} = V_k$ ; and  $ET_i$ , the end transaction.

Many transactions will be occurring concurrently and, as such, the recorded sequences of the multiple transactions will interleave. Therefore, if  $T_i$  is being executed and concurrently Transaction  $j$  is being executed,  $T_j =$  Update ( $R_{li}, R_{ji}, R_{ki}, \text{Read } (R_{li}, R_{mi}, R_{ni}, R_{oi})$ ), then the recorded sequence in the database log may appear as the following:

$$L = \{BT_i, R_{li} = V_i, R_{ji} = V_j, BT_j, R_{li,i} = V_{ii}, R_{ki} = V_k, R_{ji,j} = V_{ji}, ET_i, R_{ji,j} = V_{ji}, R_{ki,j} = V_{ki}, ET_j\}$$

The system and method of the present invention, which carries out update propagation, preferably, includes four components. These components are source log processing, target log processing, update preprocessing, and update propagation to the target database. Source log processing creates the update domain for the update propagation transaction. Target log processing creates the update domain for the target transactions based on which transactions have to be rolled back before the update transaction starts. Update processing involves the destaging from the cache memory all of the records in the update domain of the update propagation transaction and roll back all pending target transactions that overlap the domain. Update propagation to the target database involves setting the time for  $\lambda$ , the maximum time for reads, to ensure that the target transactions atomicity is not violated in the "unsafe zone" of transaction duration and propagate the update to the target database and its log.

In order for the system and method of the present invention to operate properly, it is understood that reads and writes to storage in the source database and target database are intercepted in an I/O buffer, such as common I/O buffer 220 shown in Figures 2 and 3. Common I/O buffer 230 will contain the data of the read and write operations, and the disk addresses to and from where the data was read or written. The update domains  $U$  of the update propagation  $T = (U)$  are obtained by

intercepting writes to the source database log. Similarly, the update domain  $U_i$  for the target transaction  $T_i$  is derived by intercepting all writes to the target database log.

The interception and extraction of reads (or  $R_i$ ) is more involved given that the read operations do not appear in the target database log. When read operations are intercepted as they are executed directly to the database, there is no information that is readily available to indicate the transaction to which it belongs. Since this information is not available, there can be no determination whether a particular read operation violated the atomicity equation. More specifically, a read operation cannot be associated with any other read or write operation in the same transaction.

According to the system and method of the present invention, each read operation is assigned a maximum time limit  $\lambda$ . This will mean that no read transaction may take longer than  $\lambda$  to complete.  $\lambda$  may be set to any value but preferably it will be in milliseconds from predetermined events. The exact value of  $\lambda$  may, however, depends on the particular application in which it is used. For example,  $\lambda$  may equal to 300ms in the Database2 update application.

When considering read operations in the context of the present invention, " $r$ " will belong to the transaction of some other read or update operation, if that other read or update operation occurred within the same time interval  $\lambda$  relative to " $r$ ." This applies to situations in the past and in the future relative to  $\lambda$ .

In light of read operations only being able to execute to locations, such as cache memory, there is not the possibility to intercept them and they are unnoticeable. The present invention overcomes this problem by flashing the read information from the cache and invalidating in the cache any record that belongs to the update domain of the update propagation transaction. This is realized when there is a start of the execution of an update propagation transaction  $T$  and there is the ability to track all of the target reads to the relevant records. Even using this process, it is possible to miss some of the relevant reads to relevant records before  $T$  started. This is consistent with the definitions set forth above relating to parallel transactions and atomicity since it could have happened before the start of  $T$ .

Referring the parallel transaction and atomicity definitions for this purpose, it is understood that based on these definition reads can only execute to a separate stored location such as a cache. It is done this way according to the present invention because to the second condition of the atomicity definition, namely, if  $O((R \cup U) \cap$

$U_i > 1$ , provides that all read/update operations which violate  $O((R \cup U) \cap U_i) < 2$ , occur either before T started or after T ended, then at least one read or update operation will have occur after T has started. As such, it is necessary only to track of relevant read and update records from the time T started for the interval  $\lambda$  or until T ends, whichever occurs last. What then remains is to compensate for the particular missing read operations, if they exist, that occurred before T started by setting the atomicity equation to  $O((R_i \cup U_i) \cap U) > 0$ .

According to the system and method of the present invention, all parallel target transactions  $T_i$  are rolled back to the update propagation transaction T for which  $O((R_i \cup U_i) \cap U) > 0$ . This will take place even before the transaction T starts executing. These same transactions may also be rolled back after update propagation transaction T starts. However, under these conditions, it is necessary to ensure that the atomicity rules are not violated.

The conditions under which the atomicity rule are not violated when the rollback takes place after update propagation transaction T has started executing is when the rollback process seeks to update one of the records in the update domain of transaction T and this record has already been updated by update propagation transaction T to a new value. A later rollback of transaction  $T_i$ , as a result of the update that occurred after T started, will not violate the atomicity rules since the rolled back values will reflect the updates made by update propagation transaction T.

Noting the foregoing, to properly carry out the system and method of the present invention, it is necessary to enforce the atomicity equation  $O((R \cup U) \cap U_i) < 2$  by rolling back all transactions  $T_i$  for  $O((R_{im} \cup U_i) \cap U) > 1$ . In fact, this rolling back may be considered to apply to all transaction for which  $O((R \cup U) \cap U) > 0$ .

As discussed above, the update propagation is carried out, preferably, using four components. These components are source log processing, target log processing, update preprocessing, and update propagation to the target database. Briefly, as discussed previously, source log processing creates the update domain for the update propagation transaction; target log processing creates the update domain for the target transactions based on which transactions have to be rolled back before the update transaction starts; update processing, which is the destaging from the cache memory all of the records in the update domain of the update propagation transaction and the rolling back of all pending target transactions that overlap the domain; and

update propagation to the target database. Each of these components will now be discussed in detail.

### Source Log Processing

The source log processing component of the system and method of the present invention will receive as its input parameter one log entry at a time. These log entries apply to the source database, which in the system shown in Figure 1 would be mainframe control unit/database 104 that has been referred to as DB1. The source database is where the updates originate. The source database includes a source transaction update table that is referred to as SourceTransactionUpdateTable<sub>i</sub>. This table lists all of the T<sub>i</sub>s update records and the new values for these records. The ProcessSourceLog of the source database implements the SourceTransactionUpdateTable<sub>i</sub>.

The ProcessSourceLog, as stated above, receives one input parameter. The log entries consist of the intercepted data and the entries are input to the log one at a time. The entries are data that has been intercepted by, and visible through, common I/O buffer 220 that is shown in Figure 2. Accordingly, common I/O buffer 220 will contain the successive intercepted source log data that has been input to the system. That is, I/O buffer 220 will contain the next intercepted source log I/O data. This intercepted data will provide the SourceTransactionUpdateTable<sub>i</sub> with the data it will form a list of tables for all of the open source transactions. The source log processing component will operate according to the following:

```

ProcessSourceLog(IObuffer, SourceTransactionUpdateTable)
{
    if IObuffer contains BTk then
        create new table SourceTransactionUpdateTablek
    if IObuffer contains Rj,k and its value Vj then
        add Rj and its value Vj to SourceTransactionUpdateTablek
    if IObuffer contains ETk then
        mark Source TransactionUpdateTablei as ready for update
        propagation
    if IObuffer contains ATk then /* is an abort transaction marker */
        clear and release SourceTransactionUpdateTablek
}

```

The results of the source log processing will be the listing of all of the source transactions that will be available for use by the system components.

### Target Log Processing

The second component is the target log processing component. This component is to track all of the pending target transactions that execute in parallel to the update propagation transaction. These log entries apply to the database which in the system shown in Figure 1 is open system database 108. This database has been previously referred to as DB2, the target database where the updates are to propagate.

The purpose of obtaining this target log information is to rollback all of the parallel target transactions  $T_i$  to the update propagation transaction  $T$  for which  $O(U_i \cap U) > 0$ . As stated previously, this may take place before transaction  $T$  has started executing.

Similar to the source log processing component, the input parameter to the target log processing component will be one log entry at a time. These input parameters will be included in  $\text{TargetTransactionUpdateTable}_i$ . The list will include  $T_i$  update records and their new values.

The  $\text{TargetTransactionUpdateTable}_i$  is implemented through the  $\text{ProcessTargetLog}$  that is part the target database. The input parameters are received from the common I/O buffer 220 that is shown in Figure 2. That is, the entries are data that has been intercepted by, and visible through, common I/O buffer 220. Common I/O buffer 220 will contain the successive intercepted source log data that has been input to the system. This intercepted data will provide the  $\text{TargetTransactionUpdateTable}_i$  with the data it will form a list of tables of all open transactions. The target log processing component operates according to the following:

```

ProcessTargetLog(IObuffer, TargetTransactionUpdateTable)
{
    if IObuffer contains  $BT_k$  then
        create new table  $\text{TargetTransactionUpdateTable}_k$ 
    if IObuffer contains  $r_{j,k}$  and its value  $V_j$  then
        add  $r_j$  and its value  $V_j$  to  $\text{TargetTransactionUpdateTable}_k$ 
    if IObuffer contains  $ET_k$  then
        clear and release  $\text{TransactionUpdateTable}_k$ 
    if IObuffer contains  $AT_k$  then /* is an abort transaction marker */
        clear and release  $\text{SourceTransactionUpdateTable}_k$ 
}

```

The results of the target log processing will be the listing of all of the transactions that will be available for use by the system components in update propagation.

### Update Preprocessing

5 The third component is the update preprocessing component. This component, preferably, is executed before applying the update propagation transaction that may be in a SourceTransactionUpdateTable<sub>x</sub>. When the update preprocessing component is operated, preferably, two operations will take place. The first is that the cache containing the read information will be flashed to obtain this information. Next, the  
10 flashed cache locations will have those locations invalidated that correspond to each of the records in the SourceTransactionUpdateTable<sub>x</sub>. The second is that there will be a rolling back of all open target transaction T<sub>i</sub> for those cases in which  $O(U_i \cap U) > 0$ .

The update preprocessing component preferably has two input parameters. These parameters are obtained from SourceTransactionUpdateTable<sub>x</sub> and  
15 TargetTransactionUpdateTable<sub>x</sub>. As discussed, SourceTransactionUpdateTable<sub>x</sub> will contain the list of records to be updated by update propagation transaction and TargetTransactionUpdateTable<sub>x</sub> will contain a list of target transactions that have not been committed to a transaction T. The operation of the update preprocessing component is as follows:

```

20 UpdatePreprocessing (SourceTransactionUpdateTablex,
   TargetTransactionUpdateTable)
   {
       for each record in each TargetTransactionUpdateTablei do
25         {
             If TargetTransactionUpdateTablei ∩
               SourceTransactionUpdateTablex ≠ 0 then Rollback
               Transaction(k)
           }
       for each record Ri in SourceTransactionUpdateTablex do
30         {
             Map Ri to the corresponding record ri
             Destage(ri)
           }
       }
35

```

The purpose of the update processing component is to prepare for updating the target database. This component has looked at the tables in the source and target logs and ensured that upon executing the update propagation to the target database there will not be any violation of the atomicity rules. Moreover, the update preprocessing

component will also look to ensure that the reads from the target database do not violate these rules.

### Update Propagation

The update propagation component's primary function is to propagate updates to the target database. This is accomplished by executing an update transaction T based on the values in a SourceTransactionUpdateTable<sub>i</sub>. When the update transaction T starts, it will cause a timer to be initiated after a predetermined number of  $\lambda$  units have expired. When this timer expires, it will remove the particular SourceTransactionUpdateTable<sub>i</sub> from the system. It will also be removed if the transaction is completed before the timer expires. The controlling event of the two will be the one that happens last. During the operation of the update propagation component, the system and method of the present invention will evaluate the impact of transaction T on the atomicity of any other target transactions until the SourceTransactionUpdateTable<sub>i</sub> has been removed from the system.

The update propagation component receives one parameter. This parameter is what is present in SourceTransactionUpdateTable<sub>i</sub>. These will be the lists or records to be updated and their corresponding update values. The update propagation component operates according to the following:

```

UpdateDB(SourceTransactionUpdateTablei)
{
    set timer event in T time unites for a SourceTransactionUpdateTablei
    /*propagate updates
    Create new unique transaction id (k)
    Write BTk to the log
    for each mapped record ri in SourceTransactionUpdateTablei do
    {
        read rj before-image value BVj
        write rj,k and its before-image value BVj to the target log
        write value Vj to record ri
        write rj,k and its value to the target log
    }
    Write ETk to the log
  
```

Once the update propagation process has been carried out, the target database will be update with the desired new information at the I/O stream without the need to use valuable CPU time or TCP/IP communications lines.

It is understood that in discussing the update process of the present invention that the database operations of rollback and flashing the cache may be programmed. It is also understood that the actions of rollback and flashing the cache are performed conventionally.

5 Considering the system and method as described above, and embodiment of the invention is carried out according to the following:

UpdatePropagation()

{

do forever

{

intercept next read or write operation to IObuffer  
case(IObuffer)

{

write to source log database:

/\*add record to the right TransactionUpdateTable

ProcessSourceLog(IObuffer, SourceTransactionUpdateTable)

Write record to source log /\*forward write to destination

If record is ET<sub>j</sub> then /\*start update propagation transaction

/\*spun update propagation sequence to be executed as a  
separate task

/\*parallel with intercepting reads and writes

fork (UpdateDB(TransactionUpdateTable<sub>i</sub>))

write record r<sub>j,k</sub> to target log:

ProcessTargetLog(IObuffer, TargetTransactionUpdateTable)

Write record to target log /\*Forward write to destination

/\*if write compromising transaction's consistency rollback the  
transaction

If r<sub>j,k</sub> such that r<sub>j</sub> in SourceTransactionUpdateTable then

{

RollbackTransaction(k)

clear TargetTransactionTable<sub>k</sub>

}

read r from target database

/\*if reads compromise transaction's atomicity rollback

if r such that r in some SourceTransactionUpdateTable then  
RollBackRecord(r)

/\*end case

/\*end do forever

Time event interrupt for transaction j:

If SourcePendingTransactionTable<sub>j</sub> not cleared yet and marked  
completed then

Clear SourceTransactionUpdateTable<sub>j</sub>

End of update propagation transaction task for

SourceTransactionUpdateTable<sub>j</sub>:

if time event for SourceTransactionUpdateTable<sub>j</sub> occurred already then

Clear SourceTransactionUpdateTable<sub>j</sub>

else

mark SourceTransactionUpdateTablej as completed

}

5 In another embodiment of the system and method of the present invention, the updates will not only be written to the target database but they will be written in the source database log as a new transaction. This will permit the tracking of the update events for action at a later time, such a database reconstruction or recovering.

10 The terms and expressions that are used herein are meant to be terms of expression and not limitation. And there is no intention in the use of such terms and expressions of excluding the equivalents of the features shown and described or portions thereon, it being recognized that various modifications are possible within the scope of the present invention.

**Claims:**

1. A method for updating system databases when at least source and target system databases are associated with different system platforms, the method comprising the steps of:
  - (a) intercepting write inputs to, and read inputs from a plurality of system databases as such write and read inputs are present in input/output ("I/O") streams of the system;
  - (b) inspecting the intercepted write and read inputs to determine the write and read inputs that are for updating system databases;
  - (c) listing in a source database log of the source system database the write and read inputs that are for updating the source database;
  - (d) listing in a target database log of the target system database the write and read inputs that are for updating the source database;
  - (e) writing to the source database the write and read inputs that are for updating the source database; and
  - (f) writing to the target database the write and read inputs that are for updating the source database.
2. The method as recited in claim 1, wherein write and read inputs are intercepted by an I/O buffer.
3. The method as recited in claim 2, wherein the intercepted write and read inputs are visible through the I/O buffer.
4. The method as recited in claim 1, wherein the write and read inputs that are for updating the source database are input one at a time to the source database log.
5. The method as recited in claim 1, wherein the write and read inputs that are for updating the source database are input one at a time to the target database log.
6. The method as recited in claim 1, wherein the updates are written to the target database on a non-interfering basis with other transactions affecting the target database.
7. The method as recited in claim 6, wherein when writing the updates to the target database interferes with other transactions affecting the target database, the writing update transaction will be rolled back to a non-interfering time.

8. The method as recited in claim 1, wherein the source database is operated according to first database software.
9. The method as recited in claim 8, wherein the target database is operated according to second database software.
10. A method for propagating cross-platform database updates for a system, comprising the steps of:
  - (a) intercepting a write or read that is input to an Input/Output ("I/O") buffer, with the write or read being intended for updating a source database;
  - (b) writing each write to a source database log and a target database log;
  - (c) reading each read from the target database and associating the read with a transaction listed in the source database log and the target database log;
  - (d) writing the updates from the source database log to the source database;and
  - (e) writing the updates from the target database log to the target database, according to timing that is free from interfering with other transactions affecting the target database.
11. The method as recited in claim 10, wherein the intercepted writes and reads are visible through the I/O buffer.
12. The method as recited in claim 10, wherein the writes and reads that are for updating the source database are input one at a time to the source database log.
13. The method as recited in claim 10, wherein the writes and reads that are for updating the source database are input one at a time to the target database log.
14. The method as recited in claim 10, wherein the source database is operated according to first database software.
15. The method as recited in claim 14, wherein the target database is operated according to second database software.
16. The method as recited in claim 10, wherein when writing the updates to the target database interferes with other transactions affecting the target database, the writing update transaction will be rolled back to a non-interfering time.
17. A system for propagating cross-platform database updates, comprising

a first platform that further includes,  
a first database, with the first database including a first database log for receiving update write and read information, and  
a first processor that connects to, and controls operation of the first database

a second platform that further includes,  
a second database, with the second database including a second database log for receiving update write and read information, and  
a second processor that connects to, and controls operation of the second database; and

a host that is disposed between the first processor and database and between the second processor and database, with the host further including,  
an I/O buffer that intercepts for writes and reads for updating the first and second databases,

processor means for writing the updates to the first and second databases, with the processor means controlling the writing of the updates to the second database in a non-interfering manner with other transactions affecting the second database.

18. The system as recited in claim 17, wherein the first database is operated according to first database software.

19. The system as recited in claim 18, wherein the second database is operated according to second database software.

1/3

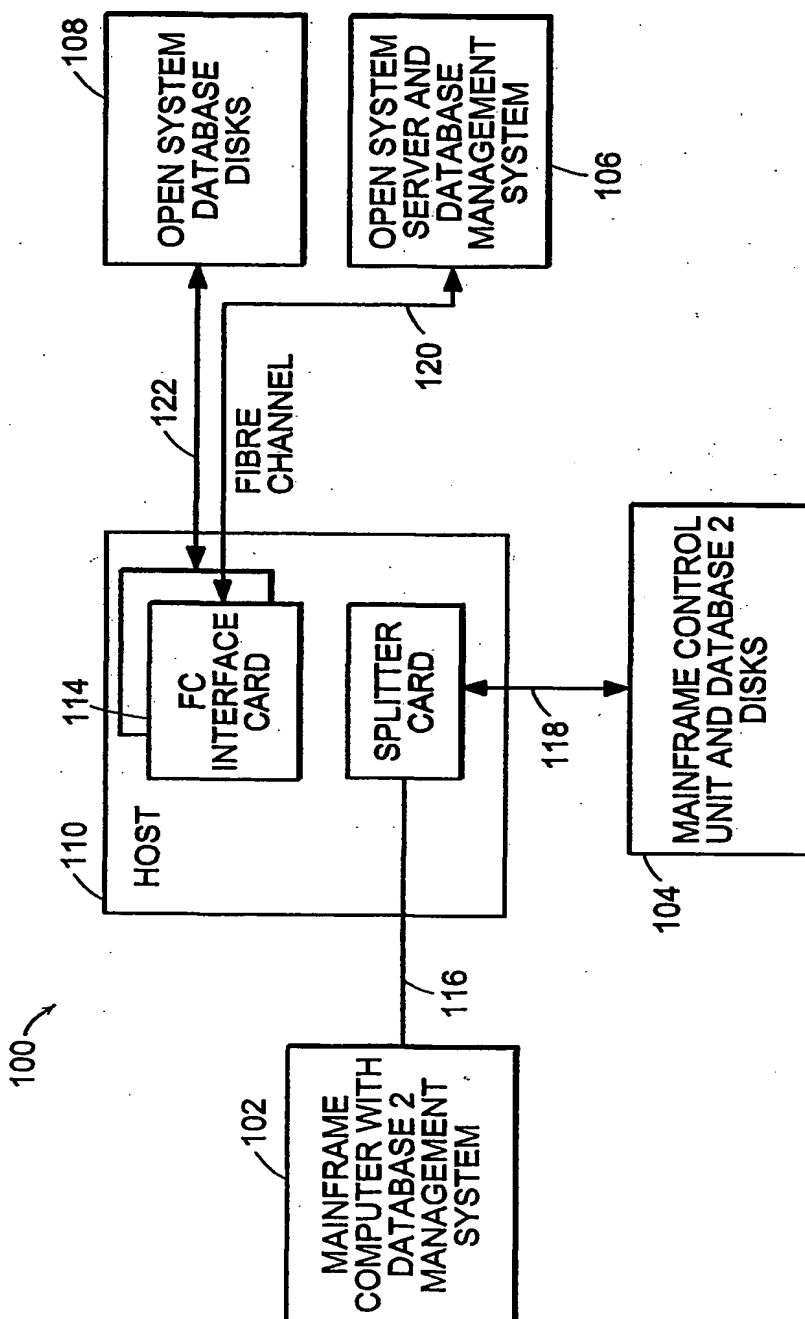


FIG. 1

2/3

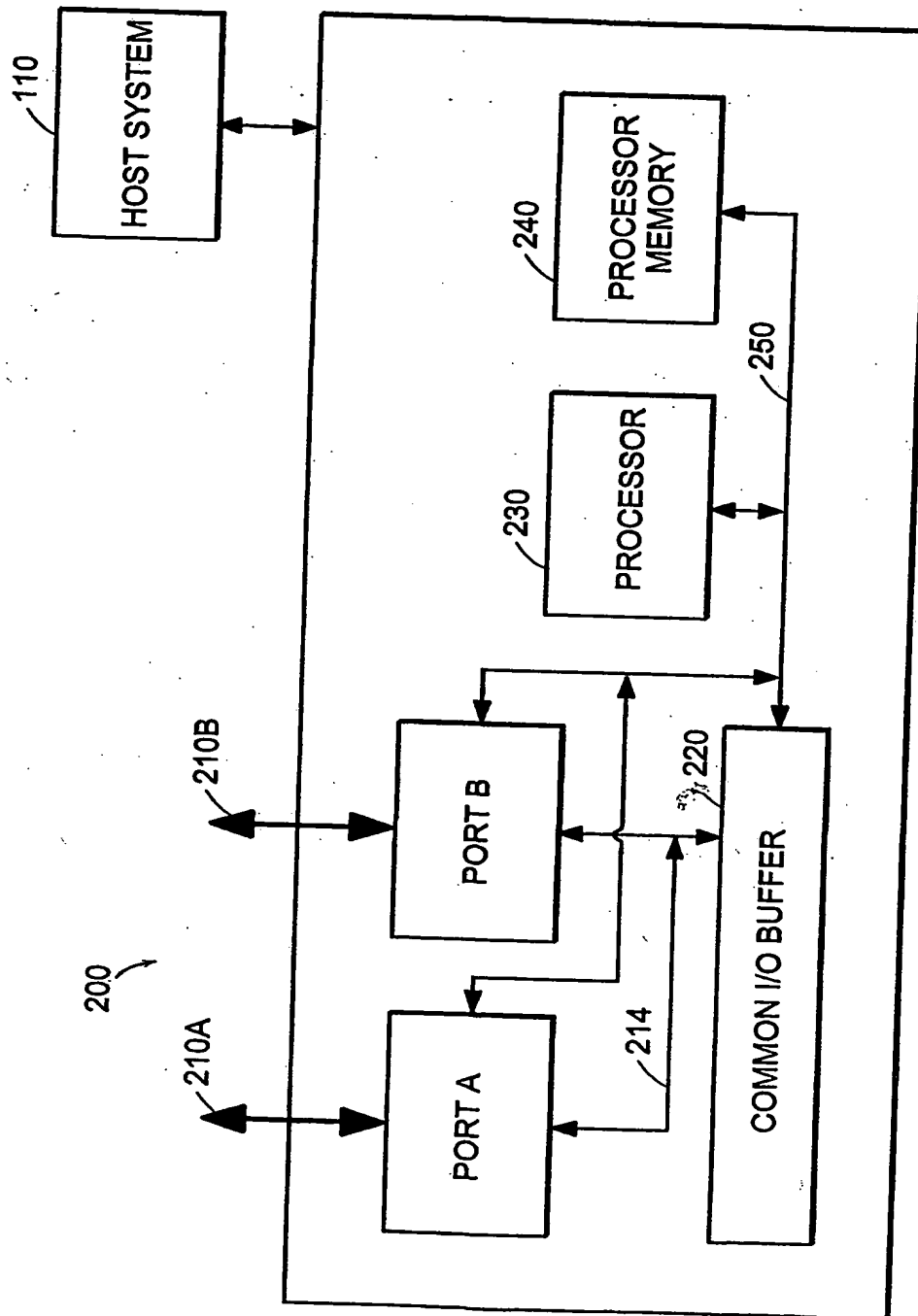


FIG. 2

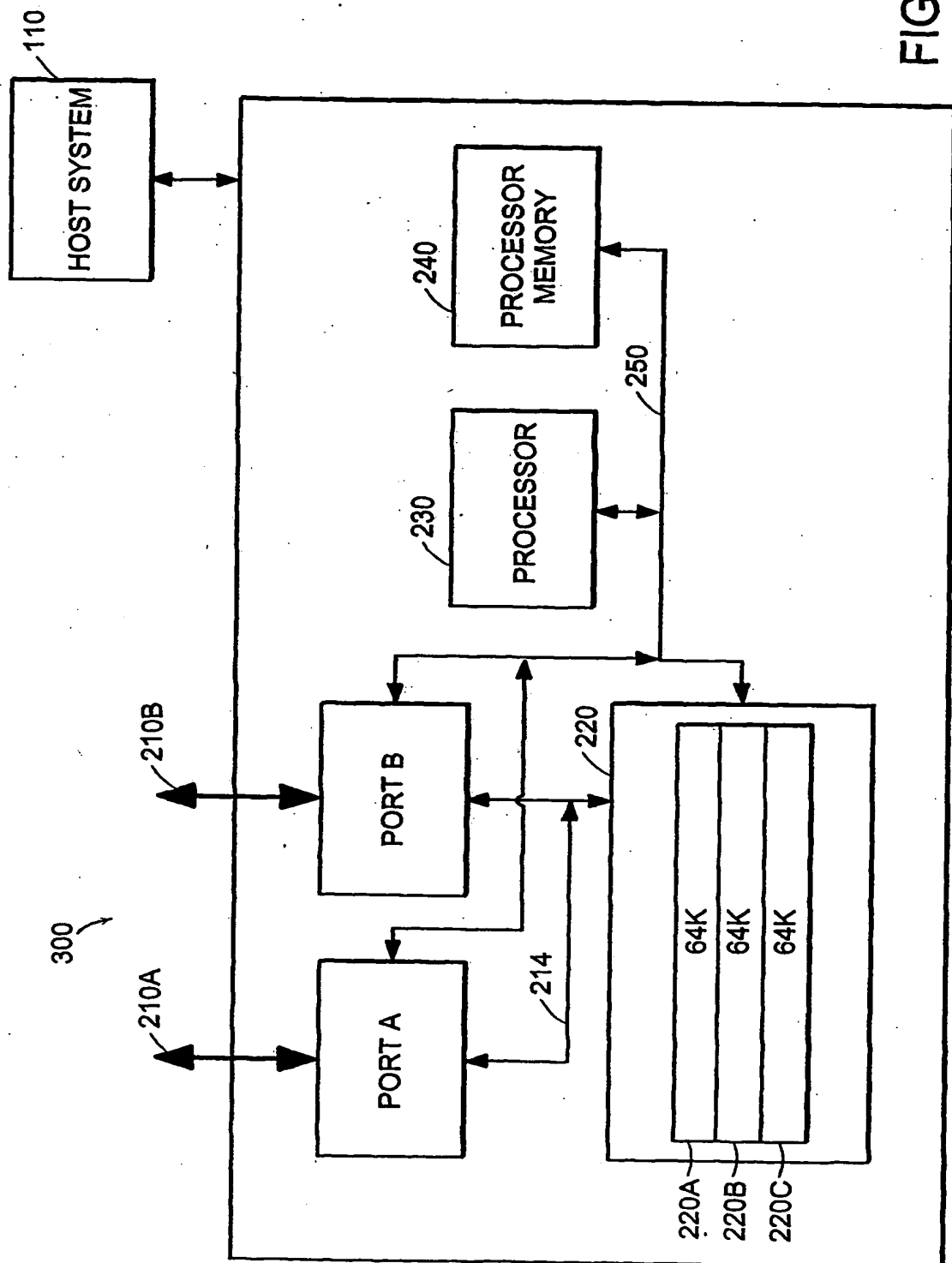


FIG. 3

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US02/01429

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :G06F 17/30

US CL :707/201

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 707/201, 8, 10, 202; 709/217, 219, 224

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
**NONE**Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
**EAST**

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,799,141 A (GALIPEAU et al) 25 August 1998, Abstract, col. 2, lines 5-46	1-19
Y,P	US 6,256,675 B1 ( RABINOVICH) 03 July 2001, Abstract, fig.1, col. 4, lines 40 to col. 5, lines 28, and col. 6, lines 6-59	1-19
Y	US 5,991,771 A (FALLS et al) 23 November 1999, Abstract, col. 3, lines 36-65, col. 4, lines 40-58, and col. 14, lines 26-67 to col. 15, lines 4-54.	1-19
A	US 6,144,999 A (KHALIDI et al) 07 November 2000, See the whole reference.	1-19
A	US 6,014,686 A (ELNOZAHY et al) 11 January 2000, See the whole reference.	1-19

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier document published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"A" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search  
22 MARCH 2002

Date of mailing of the international search report

18 APR 2002

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231Authorized officer *JM*

JACQUES VEILLARD

Telephone No. *James R. Matthews*  
(202) 305-3600